

IN THE CLAIMS:

1. (Currently Amended) A method for service time analysis in a computer network, comprising:

receiving a request from a network client machine for an electronic document and recording an initial time value for the request;

forwarding the request to an origin server and receiving a response stream containing the electronic document from the origin server;

sending the response stream to the client machine, wherein the response stream is instrumented to contain the initial time value;

receiving at least one [[a]] uniform resource identifier (URI) request from the client machine, wherein [[the]] each received URI request is for a resource embedded within the electronic document; ~~and recording a service time value for completing the request for the electronic document~~;

creating a data structure that contains key/service time value pairs, wherein each pair is associated with one of said URI requests, and each key is an identifier present in its associated URI request;

forwarding [[the]] each URI request to the origin server, and receiving a corresponding URI response from the origin server;

updating each of the service time values; ~~and in said data structure, to indicate the time value for completing the response to each of said URI requests~~;

sending [[the]] each URI response to the client machine~~[[.]]~~; and

using said keys and updated service time values from said data structure to compute the total time required to respond to all of said requests associated with said electronic document, including all of said URI requests.

2. (Currently Amended) The method according to claim 1, further comprising a ~~wherein said data structure comprises a table, and each key in said table of key/time value table, wherein the key pairs is a cookie in a request header, and the value is a time stamp signifying the service time for a request~~.

said table of key/time value pairs is created and updated by at least one reverse proxy server associated with said origin server.

3. (Currently Amended) The method according to claim 1, wherein the instrumented response stream further comprises:

~~the initial time value of the request; and~~

the service time taken for the origin server to respond.

4. (Original) The method according to claim 1, wherein all steps are performed by a single reverse proxy server.

5. (Currently Amended) The method according to claim 1, wherein the steps are performed by multiple reverse proxy servers, wherein a selected one of said reverse proxy servers acts as the controlling quality-of-service monitor, and the other remaining reverse proxy servers are subordinate servers to the controlling monitor.

6. (Original) The method according to claim 5, wherein:

the controlling monitor sends a sample-on command to the subordinate servers;

in response to the sample-on command, the subordinate servers record service time metrics for request transactions;

the controlling monitor sends a sample-off command to the subordinate servers;

in response to the sample-off command, the subordinate servers send their respective service time records to the controlling monitor; and

the controlling monitor analyzes and reorganizes the service time records from the subordinate servers into a single record.

7. (Currently Amended) A computer program product in a computer readable medium for use in a data processing system, for service time analysis in a computer network, the computer program product comprising:

instructions for receiving a request from a network client machine for an electronic document and recording an initial time value for the request;

instructions for forwarding the request to an origin server and receiving a response stream containing the electronic document from the origin server;

instructions for sending the response stream to the client machine, wherein the response stream is instrumented to contain the initial time value;

instructions for receiving [[a]] at least one uniform resource identifier (URI) request from the client machine, wherein [[the]] each received URI request is for a resource embedded within the electronic document, and recording a service time value for completing the request for the electronic document;

instructions for creating a data structure that contains key/service time value pairs, wherein each pair is associated with one of said URI requests, and each key is an identifier present in its associated URI request;

instructions for forwarding [[the]] each URI request to the origin server, and receiving a corresponding URI response from the origin server;

instructions for updating each of the service time values, and in said data structure, to indicate the time value for completing the response to each of said URI requests;

instructions for sending [[the]] each URI response to the client machine[[.]]; and instructions for using said keys and said updated service time values from said data structure to compute the total time required to respond to all of said requests associated with said electronic document, including all of said URI requests.

8. (Currently Amended) The computer program product according to claim 7, further comprising wherein said data structure comprises a table, and each key in said table of key/time value table, wherein the key pairs is a cookie in a request header, and the value is a time stamp signifying the service time for a request.

said table of key/time value pairs is created and updated by at least one reverse proxy server associated with said origin server.

9. (Currently Amended) The computer program product according to claim 7, wherein the instrumented response stream further comprises:

the initial time value of the request, and

the service time taken for the origin server to respond.

10. (Original) The computer program product according to claim 7, wherein all instructions are performed by a single reverse proxy server.

11. (Currently Amended) The computer program product according to claim 7, wherein the instructions are performed by multiple reverse proxy servers, wherein a selected one of said reverse proxy servers acts as the controlling quality-of-service monitor, and the other remaining reverse proxy servers are subordinate servers to the controlling monitor.

12. (Original) The computer program product according to claim 11, further comprising:

instructions for sending a sample-on command from the controlling monitor to the subordinate servers;

in response to the sample-on command, instructions for the subordinate servers to record service time metrics for request transactions;

instructions for sending a sample-off command from the controlling monitor to the subordinate servers;

in response to the sample-off command, instructions for the subordinate servers to send their respective service time records to the controlling monitor; and

instructions for the controlling monitor to analyze and reorganize the service time records from the subordinate servers into a single record.

13. (Currently Amended) A system for service time analysis in a computer network, comprising:

a first receiving component which receives a request from a network client machine for an electronic document and records an initial time value for the request;

a first communication component which forwards the request to an origin server and receives a response stream containing the electronic document from the origin server;

a second communication component which sends the response stream to the client machine, wherein the response stream is instrumented to contain the initial time value;

a second receiving component which receives [[a]] at least one uniform resource identifier (URI) request from the client machine, wherein the each received URI request is for a resource embedded in the electronic document and records a service time value for completing the request for the electronic document;

a component for creating a data structure that contains key/service time value pairs, wherein each pair is associated with one of said URI requests, and each key is an identifier present in its associated URI request;

a third communication component which forwards [[the]] each URI request to the origin server, and receives a corresponding URI response from the origin server;

an updating component which updates each of the service time values, and in said data structure, to indicate the time value for completing the response to each of said URI requests;

a fourth communication component which sends [[the]] each URI response to the client machine[[.]]; and

a processing component adapted to use said keys and said updated service time values from said data structure to compute the total time required to respond to all of said requests associated with said electronic document, including all of said URI requests.

14. (Currently Amended) The system according to claim 13, wherein said data structure comprises a table and said system further comprising comprises a register which maintains [[a]] said table of key/time value table pairs, wherein [[the]] each key is a cookie in a request header, and the value is a time stamp signifying the service time for a request.

said table of key/time value pairs is created and updated by at least one reverse proxy server associated with said origin server.

15. (Currently Amended) The system according to claim 13, wherein the instrumented response stream further comprises:

the initial time value of the request; and

the service time taken for the origin server to respond.

16. (Original) The system according to claim 13, wherein all components are contained in a single reverse proxy server.

17. (Currently Amended) The system according to claim 13, wherein the components are contained in multiple reverse proxy servers, wherein a selected one of said reverse proxy servers acts as the controlling quality-of-service monitor, and the other remaining reverse proxy servers are subordinate servers to the controlling monitor.

18. (Original) The system according to claim 17, further comprising:

a first communication component which sends a sample-on command from the controlling monitor to the subordinate servers;

a plurality of recording components in the subordinate servers which, in response to the sample-on command, record service time metrics for request transactions;

a second communication component which sends a sample-off command from the controlling monitor to the subordinate servers;

a plurality of response components in the subordinate servers which, in response to the sample-off command, send their respective service time records to the controlling monitor; and

a processor in the controlling monitor which analyzes and reorganizes the service time records from the subordinate servers into a single record.